yang_seonhyeHW13

Seonhye Yang

2/19/2019

```
install.packages("car")
```

```
## Installing package into '/home/rstudio-user/R/x86_64-pc-linux-gnu-library/3.5'
## (as 'lib' is unspecified)
```

```
library(car)
```

```
## Loading required package: carData
```

```
library(alr4)
```

```
## Loading required package: effects
```

```
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
```

Question 1

```
scottish_recycle <- read.fwf("http://users.stat.ufl.edu/~winner/data/scottish_recycle.dat", c(27, 11, 1
                             col.names = c("Local Authority", "Recycling Capacity", "Residual Capacity"
attach(scottish_recycle, warn.conflicts = FALSE)
model <- lm(Yield ~ Recycling.Capacity + Residual.Capacity + Extended.Materials)

vif(model)
```

```
## Recycling.Capacity  Residual.Capacity Extended.Materials
##           2.056995           1.653963           1.417643
```

None of our variables have a VIF of more than 10, so multicolinearity is not suggested.

Problem 2

```
hotel.energy <- read.csv("http://www.stat.ufl.edu/~winner/data/hotel_energy.csv")
attach(hotel.energy, warn.conflicts = FALSE)
model1 <- lm(enrgcons ~ area + age + numrooms + occrate)

vif(model1)
```

```
##     area      age numrooms  occrate
## 3.757655 1.264324 3.782078 1.224289
```

Again, none of our variables have a VIF of more than 10, so multicolinearity is not suggested.

Question 3

```
cruise_data <- read.fwf("http://users.stat.ufl.edu/~winner/data/cruise_ship.dat", c(20, 26, 4, 9, 8, 8,
                        col.names = c("Ship Name", "Cruise Line", "Age", "Tonnage", "Passengers", "Leng
attach(cruise_data, warn.conflicts = FALSE)
model2 <- lm(Passengers ~ Age + Tonnage + Length)

vif(model2)
```

```
##              GVIF Df GVIF^(1/(2*Df))
## Age      8.830203 37        1.029872
## Tonnage 13.586201  1        3.685946
## Length  10.277880  1        3.205913
```

The variables Tonnage and Length have VIF greater than 10, so multicolinearity is strongly suggested, indicating a problem.

Question 4

```
data("MinnWater")
attach(MinnWater, warn.conflicts = FALSE)
model3 <- lm(muniUse ~ ., data=MinnWater)

vif(model3)
```

```
##       year     allUse     irrUse   agPrecip muniPrecip   statePop
##  633.34563  190.15277  118.11767    5.72228    4.28763 1904.44626
##    muniPop
## 3441.37710
```

The variables year, allUse, irrUse, statePop, and muniPop all have VIFs greater than 10, suggesting multicolinearity. The largest of these is muniPop, with a value of 3441.377. This indicated that muniPop is a strongly redundant variable in our model and can be removed to improve the standard errors of our coefficients, making our model better.

Question 5

We know that the formula for VIF is $\texttt{VIF} = \frac{1}{1-R^2_{X_i}}$, and we are given the fact that $\texttt{VIF}_{X_1} = 11.2$, so we can solve accordingly:

$$11.2 = \frac{1}{1 - R^2_{X_1}}$$

$$1 - R^2_{X_1} = \frac{1}{11.2}$$

$$R^2_{X_1} = 1 - \frac{1}{11.2}$$

$$R^2_{X_1} = 0.911$$

We can then interpret this to mean that $X_2 \dots X_5$ account for 91.1% of X_1's variance.\

Multicolinearity is important to check in any model because the fact that a variable might be strongly colinear with another independent variable could lead to large coefficients that would make variables that should not be significant into significant ones. So when we see a variable with a large VIF, we should think about removing it from our model "'